UNIVERSITÉ DE MONTRÉAL

AN INTRODUCTION TO NETWORK MODELS USED IN TRANSPORTATION PLANNING

by

Michael Florian

CENTRE DE RECHERCHE SUR LES TRANSPORTS

CASE POSTALE 6128, SUCCURSALE A, MONTRÉAL, QUÉBEC, H3C 3J7

AN INTRODUCTION TO NETWORK MODELS USED IN TRANSPORTATION PLANNING

by

Michael Florian

This paper was written to serve as notes for the First Course on Transportation Planning Models of the International School of Transport Planning given at Amalfi, Italy, October 11-16, 1982. The underlying research was supported in part by a grant from the National Sciences and Engineering Research Council Canada. Heinz Spiess made some helpful suggestions regarding the text.

Université de Montréal

Centre de recherche sur les transports - Publication #262
Département d'informatique et de
recherche opérationnelle - Publication #448

Août 1982

Dépôt légal - 3e trimestre 1982

ABSTRACT

The user optimal and systems optimal models or route choice in transportation networks are formulated and their mathematical properties are described. Solution algorithms for fixed and variable demand versions of this model are studied. Network models with constant travel costs are presented as well. The application of these models in various practical contexts is described and qualified.

THE INTEREST IN THE MATHEMATICAL MODELING OF TRANSPORTATION SYSTEMS stems from the need to predict how people, vehicles and goods might make use of new or improved transport infrastructure in order to evaluate the benefits of the required investments. To this end it is necessary to build models of the demand for transportation and models that determine the way in which people, vehicles and goods use the transportation network. If such models may be constructed and their validity reasonably assured, then the prediction of the traffic flows on present and future transportation links may be carried out by simulating future situations and then evaluating the potential benefits of alternative improvement projects.

In the attempts that were made to construct mathematical models of route choice in transportation networks, the notion of equilibrium plays a central role. Suppose that the demand is known between an origin and a destination. Then it is of central importance to determine how these trips are attracted to the alternative routes available between origins and destinations. The idea of traffic equilibrium goes back to at least 1924, when Knight [22] gave a simple and intuitively clear description of a postulate of traffic behavior under conditions of congestion. We quote :

"Suppose that between two points there are tow highways, one of which is broad enough to accomodate without crowding all the traffic which may care to use it, but is poorly graded and surfaced, while the other is a much better road, but narrow and quite limited in capacity. If a large number of trucks operate between the two termini and are free to choose either of the two routes, they will tend to distribute themselves between the roads in such proportions that the cost per unit of transportation, or effective returns per unit of investment, will be the same for every truck on both routes. As more trucks use the narrower and better road, congestion develops, until a certain point it becomes equally profitable to use the broader but poorer highway."

Some twenty eight years later, Wardrop [34] stated the two principles that formalized this notion of equilibrium and introduced the alternative behavior postulate of the minimization of total costs. His first principle states that "the journey times on all routes actually used are equal and less than those which would be experienced by a single vehicle on any unused route". The implication of this principles is that the routes actually used are the shortest under the prevailing traffic conditions and their perception by the user. This characterization of equilibrium of route choice, or traffic equilibrium, which is identical to the notion described by Knight, became accepted over the past 30 years as a sound behavioral principle to describe the subdivision of trips over alternative routes. The traffic flows that satisfy this principle are usually refered to as "user optimized" flows, since each user chooses the route that he perceives to be the best. By contrast, the "system optimized" flows are characterized by Wardrop's second principle which states that "the average journey time is minimum". Wardrop compared these two principles and commented that the first criterion is quite a likely one in practice, since it implies an equilibrium situation in which no driver can reduce his journey time by choosing an alternate route, the second criterion is the most efficient since it minimizes the total travel time.

Although Wardrop discussed the equilibrium conditions for a general network, he did not propose a method to compute the corresponding flows. Beckmann, McGuire and Winsten [2] were the first to formulate and analyze a mathematical model of traffic equilibrium on a network. Their seminal work, which appeared in 1956, was the starting point for the contributions to follow in this area which were achieved mostly within the past fifteen years.

The purpose of these lectures is to present the elements of the network models used in transportation planning, review their structural properties and most commonly used solution methods and outline current and potential applications.

THE FORMULATION OF NETWORK MODELS

The network models that are most commonly used are steady state models, in spite of the fact that almost all traffic phenomena are temporal. It is common to consider a given period of time, determine the demand characteristics during that period and determine the flow pattern that results from the interaction of the demand and the congestion characteristics of the transport infrastructure available.

Consider a transportation network which permits the flow of one type of traffic (vehicles or passengers) on its links. The nodes n, $n \in N$, represent origins, destinations and intersections of links, the links a, $a \in A$, represent the transportation infrastructure. The flow of trips on link a is v_a and the cost of travelling on a link is given by a user cost function $s_a(v)$, where v is the vector of link flows over the entire network. This cost function may model the time delay for travel on that arc, in which case it is commonly refered to as a volume/delay function, however it may model other costs, such as fuel consumption. The vector user cost functions s(v) is assumed to be monotone

$$(s(v') - s(v''))^T (v' - v'') \geq 0, \qquad a \in A \tag{1}$$

continuous and differentiable.

The origin to destination demands, g_i, $i \in I$, where I is the set of origin/destination (O/D) pairs, may use directed paths k, $k \in K_i$, where K_i is the set of paths for O/D pair i ($K_i \neq \phi$). The flows on paths k, h_k, satisfy conservation of flow and nonnegativity

$$\left.\begin{aligned} &\sum_{k \in K_i} h_k = g_i, \quad i \in I \\ &\text{and } h_k \geq 0, \quad k \in K_i, \ i \in I \end{aligned}\right\} \tag{2}$$

The corresponding link flows v_a are given by

$$v_a = \sum_{i \in I} \sum_{k \in K_i} \delta_{ak} \cdot h_k, \qquad a \in A \tag{3}$$

where $\delta_{ak} = \begin{cases} 1 & \text{if link a belongs to path k} \\ 0 & \text{otherwise} \end{cases}$

The cost of each path $s_k (= s_k(v))$ is the sum of the user costs of the links in the path

$$s_k = \sum_a \delta_{ak} \cdot s_a(v), \quad k \epsilon K_i, \; i \epsilon I \tag{4}$$

Let $u_i (= u_i(v))$ be the cost of the least cost path for any O/D pair i

$$u_i = \min_{k \epsilon K_i} s_k, \quad i \epsilon I \tag{5}$$

The demands g_i are given by functions $G_i(u)$, where u is the vector of least cost travel times for all the O/D pairs of the network

$$g_i = G_i(u), \quad i \epsilon I \tag{6}$$

The vector demand functions G.(u) is assumed to be strictly antitone

$$(G(u') - G(u''))^T (u_i' - u_i'') < 0, \quad i \epsilon I \tag{7}$$

continuous, differentiable and bounded from above.

The network equilibrium model is formulated by supposing that for every O/D pair, Wardrop's user optimal principle is satisfied, that is

$$s_k^* - u_i^* \begin{cases} = 0 & \text{if } h_k^* > 0 \\ \geq 0 & \text{if } h_k^* = 0 \end{cases} \qquad k \epsilon K_i, \; i \epsilon I \tag{8}$$

over the feasible set (2)-(3).

It is relatively straightforward to show that (8) may be restated in the "complementarity" form

$$u_i^* \leq s_k^* \text{ and } (s_k^* - u_i^*) h_k^* = 0, \quad k \epsilon K_i, \; i \epsilon I \tag{9}$$

and that (8) and (5) are equivalent to

$$s_{k_1}^* \leq s_{k_2}^*, \quad \text{if } h_{k_1}^* > 0, \; k_1, k_2 \epsilon K_i, \; i \epsilon I \tag{10}$$

Thus, Wardrop's first principle may be stated mathematically in several forms. However, the most useful restatement of the network equilibrium model is in the form of a variational inequality (Smith [32], Dafermos [7,8]). This accomplished by noting that (8) is equivalent to

$$(s_k^* - u_i^*)(h_k - h_k^*) \geq 0, \quad k \epsilon K_i, \; i \epsilon I \tag{11}$$

where h_k is any feasible path flows. If $h_k^* > 0$ then $s_k^* = u_i^*$ since h_k may be smaller than h_k^*. If $h_k^* = 0$, then (11) is satisfied when $s_k^* - u_i^* \geq 0$. By summing over $k \epsilon K_i$ and $i \epsilon I$

$$\sum_{i \epsilon I} \sum_{k \epsilon K_i} s_k^* (h_k - h_k^*) \geq \sum_{i \epsilon I} u_i^* (g_i - g_i^*) \tag{12}$$

Using (3) and (4) and by changing order of summation on the left hand side it follows that

$$\sum_{a\in A} s_a(v^*)(v_a - v_a^*) \geq \sum_{i\in I} u_i^*(g_i - g_i^*) \qquad (13)$$

Since the vector demand function is antitone, it is invertible. Let $W(g)=u$ be the inverse of the demand function. By making the corresponding substitution for u, the network equilibrium model may be stated as

$$\sum_{a\in A} s_a(v^*)(v_a - v_a^*) - \sum_{i\in I} W_i(g^*)(g_i - g_i^*) \geq 0 \qquad (14)$$

over the feasible set (2)-(3).

It may be verified that (14) implies (8) by constructing a flow pattern which differs from the equilibrium flows on only one path k', $k'\in K_i$, for which $h_{k'} = h_{k'}^* + \delta$, $0 \leq |\delta| < h_{k'}^*$. Applying (14) yields $s_{k'}^*\cdot\delta - W_i(g^*)\delta = 0$ which is equivalent to (8).

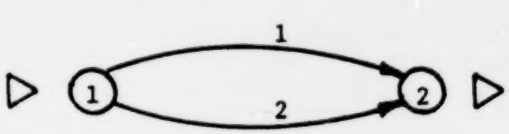
(14) may be rewritten by using matrix notation as

$$s(v^*)^T(v - v^*) - W(g^*)^T(g - g^*) \geq 0 \qquad (15)$$

In the next section, a review of some mathematical properties of the network equilibrium model is presented.

SOME PROPERTIES OF THE NETWORK EQUILIBRIUM MODEL

The existence of a solution of the network equilibrium model is ensured by the continuity of the user cost and demand functions and by the fact that the feasible set (2)-(3) is compact if cycle flows do not occur and the demand functions are bounded from above (see [1,21]). The following example illustrates that, when the user cost functions are not continuous, flows that satisfy the equilibrium conditions do not exist. The network consists of one O/D pair and two links as shown below.



$$s_1(v_1) = 4v_1$$

$$s_2(v_2) = \begin{cases} 4v_2, & v_2 < 20 \\ 4v_2 + 10, & v_2 \geq 20 \end{cases}$$

The demand for 1 to 2 is 40. When $v_1 = v_2 = 20$, the cost of link 2 is higher than the cost of link 1, but for $v_1 = 20 + \varepsilon$ and $v_2 = 20 - \varepsilon$, $\varepsilon > 0$, the cost of link 1 is higher than the cost of link 2.

In practice, the continuity requirement is not a hindrance, and in fact, all user cost functions that have been calibrated are continuous. The same is true of demand functions used in practice.

An important property of a network model that is used to simulate traffic flows for different future network configurations, is that for a given scenario the flows and origin to destination costs be unique. Unless this were so, the comparison of the different future situations would be difficult to carry out; differences between scenarios would depend on which solution was found for each case as well as on the different network configurations. Fortunately, for many situations of interest, the network equilibrium model has unique flows, demand and origin to destination costs. This is ensured when the user cost functions are strictly monotone and the

demand functions (and their inverses) are strictly antitone, as assumed in the model formulation.

To demonstrate this (see Smith [22], Dafermos [8], Aashtiani and Magnanti [1]), suppose that there exist two distinct solutions (v',g') and (v'',g''). By writing (15) once with $v=v'$, $g=g'$ and $v^*=v''$, $g^*=g''$ and once with $v=v''$, $g=g''$ and $v^*=v'$, $g^*=g'$, one obtains by adding the two inequalities

$$(s(v')-s(v''))^T(v'-v'') - (W(g')-W(g''))^T(g'-g'') \leq 0 \qquad (16)$$

However by (1) and (7) it follows that (16) is satisfies if and only if each term is equal to zero. Thus

$$(s(v')-s(v''))^T(v'-v'') = 0, \quad a\epsilon A \qquad (17)$$

and

$$(W(g')-W(g''))^T(g'-g'') = 0, \quad i\epsilon I \qquad (18)$$

It follows then from (18) that $g'=g''$, and from (17) that $s_a(v')=s_a(v'')$, hence the link user costs are unique as well. Since u_i^*, $i\epsilon I$ are the lengths of shortest paths computed for each O/D pair i based on the user costs $s_a(v^*)$, it follows that they are unique as well. We note that if the user cost functions are strictly monotone, then the link flows are unique as well.

It is worthwhile to note that the path flows h_k are not unique, in general. Given the equilibrium link flows, the corresponding path flows are given as the solution of the simultaneous linear equations

$$v_a^* = \sum_{i\epsilon I} \sum_{k\epsilon K_i} \delta_{ak} \cdot h_k^*, \quad a\epsilon A \qquad (19)$$

Usually, the number of paths used is much larger than the number of links of the network, and in (19) the number of variables h_k^* far exceeds the number of constraints, which equals to $|A|$. The consequence of this fact is that, in the analysis of particular network configurations, the paths used may be obtained, however they do not represent reliable information, since they are not necessarily unique.

Another important property of the network equilibrium model is that it is stable. The equilibrium traffic flows depend continuously upon the travel demands and travel cost functions. Roughly speaking, small changes in the travel demands or in the user cost functions result in small changes in the traffic flows. This was demonstrated recently by Dafermos and Nagurney [6] for the general equilibrium model presented here and by Hall [20] for a version of the network equilibrium model that is discussed in the next section. This property is very desirable in applications, provided that the model is a suitable representation of the observed route choice behavior.

SOLVING FOR THE EQUILIBRIUM FLOWS

Most of the applications of the network equilibrium model in practice have been achieved for simpler versions of the model (15) subject to (2)-(3). The user cost functions that have been calibrated are functions of only the flow in the link, that is

$s_a(v) = s_a(v_a)$, $a \epsilon A$ and the demand functions depend only on the travel cost between the origin and the destination of the node pair i, that is $G_i(u) = G_i(u_i)$, $i \epsilon I$. The assumptions that were made on the function properties may restated simpler to be a requirement that the user cost functions are all continuous and increasing and that the demand functions are continuous, strictly decreasing and bounded from above. The variational inequality (14) may now be restated as

$$\sum_{a \epsilon A} s_a(v_a^*)(v_a - v_a^*) - \sum_{i \epsilon I} W_i(g_i^*)(g_i - g_i^*) \geq 0 \qquad (20)$$

over the set (2)-(3).

The application of this model in practice was facilitated by the equivalence between this variational inequality and a convex cost optimization problem for which the optimality condition is (20). This optimization problem in

$$\text{Min } Z = \sum_{a \epsilon A} \int_0^{v_a} s_a(x)dx - \sum_{i \epsilon I} \int_0^{g_i} W_i(y)dy \qquad (21)$$

subject to (2)-(3).

Since the user cost functions are increasing and the demand functions are strictly decreasing, the objective (21) is convex and (20) is interpreted as the first order optimality condition, that the directional derivative of Z is larger than or equal to zero (see Property 6 of Appendix I). A rigorous proof of this equivalence may be found in Kinderlehrer and Stampacchia [21]. This is the classical result found by Beckmann *et al.* [2]. The Kuhn-Tucker conditions of the convex cost optimization problem (21) subject to (2)-(3) (see Appendix III) are equivalent to the statement of Wardrop's user optimal principle of route choice in the form (8).

The existence of the equivalent optimization problem greatly facilitated the implementation of the network equilibrium model in practice, since a number of efficient algorithms are available for solving such problems, while the solution of variational inequalities of large scale is more difficult. It may be said without reservation that most of the practical applications of the network equilibrium model are based on the initial work of Beckmann *et al.* even though recently, the variational inequality formulation opened the way to understanding and solving network equilibrium models of increasing complexity.

From the large repertory of methods that are available for solving convex cost minimization problems subject to linear constraints, the adaptation of the linear approximation method of Frank and Wolfe [18] and some of its variants is the most commonly used algorithm. Bruynooghe, Gibert and Sakarovitch [4] were the first to propose the method, however the subsequent work of LeBlanc, Morlok and Pierskalla [24] and Nguyen [29] made this method popular in practice. (See also [25,30] and [9] for a different approach). Computer codes are widely available for solving the fixed demand version of this problem, that is when $g_i = \bar{g}_i$, $i \epsilon I$. We proceed to present the details of the adaptation of this method to solve the fixed demand network equilibrium problem and then discuss some of its variants, both for the fixed and variable demand versions.

We consider now the fixed demand network equilibrium problem

$$\text{Min } Z = \sum_{a\epsilon A} \int_0^{v_a} s_a(x)dx \tag{22}$$

$$\text{subject to} \quad \sum_{k\epsilon K_i} h_k = \bar{g}_i, \quad i\epsilon I \tag{23}$$

$$h_k \geq 0, \quad k\epsilon K_i, \quad i\epsilon I \tag{24}$$

and we recall (3), that is $v_a = \sum_{i\epsilon I} \sum_{k\epsilon K_i} \delta_{ak} \cdot h_k, \quad a\epsilon A$.

As described in Appendix II, starting from an initial feasible solution the linear approximation method obtains a feasible direction by linearizing the objective function, solving a linear programming subproblem and then finding an improved solution on the line segment between the current solution and the solution of the subproblem. For the network equilibrium problem the linearized subproblem at iteration ℓ is

$$\text{Min} \sum_{a\epsilon A} s_a(v_a^{\ell}) \sum_{i\epsilon I} \sum_{k\epsilon K_i} \delta_{ak} \cdot f_k \tag{25}$$

$$\text{subject to} \quad \sum_{k\epsilon K_i} f_k = \bar{g}_i, \quad i\epsilon I \tag{26}$$

$$f_k \geq 0, \quad k\epsilon K_i, \quad i\epsilon I \ . \tag{27}$$

By changing the order of summation in (25) and by using (4), the objective becomes

$$\text{Min} \sum_{i\epsilon I} \sum_{k\epsilon K_i} s_k^{\ell} \cdot f_k \tag{28}$$

The solution of this problem is obtained trivially by computing shortest paths for each O/D pair i and allocating the demand $\bar{g}_i$ to that path. Note that the complementarity slackness conditions for the dual constraints of this linear program are

$$f_k(u_i - s_k^{\ell}) = 0, \quad k\epsilon K_i \ , \quad i\epsilon I \tag{29}$$

and the dual constraints are

$$u_i \leq s_k^{\ell}, \qquad k\epsilon K_i \ , \quad i\epsilon I \tag{30}$$

The optimal u_i are the length of the shortest paths based on the current link user costs $s_a(v_a^{\ell})$ and (29) is a statement of Wardrop's user optimal principal when the link costs are constant.

The solution to the problem (25)-(27) is then given by

$$f_a^{\ell} = \sum_{i\epsilon I} \sum_{k\epsilon K_i} \delta_{ak} \, f_k \tag{31}$$

and the direction of descent is simply $d_a^{\ell} = (f_a^{\ell} - v_a^{\ell})$.

The shortest paths calculations required in the solution of

this subproblem may be carried out by any of the efficient methods currently available.

An iteration of the linear approximation algorithm is completed by finding the solution of

$$\underset{0\le\lambda\le 1}{\text{Min}} \sum_a \int_0^{v_a^\ell+\lambda d_a^\ell} s_a(x)dx) \qquad (32)$$

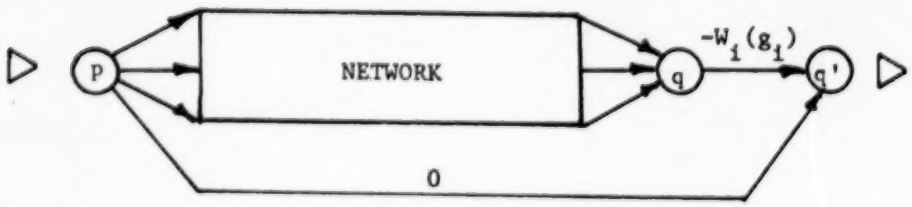
or equivalently annuling its gradient, that is find λ^ℓ for which

$$\sum_{a\in A} s_a(v_a^\ell+\lambda^\ell d_a^\ell)\, d_a^\ell = 0 \qquad (33)$$

For numerical reasons it is preferable to find λ by (33). The linearization is repeated for $\ell = 1,2,\ldots$ until a satisfactory solution is obtained.

This algorithm has several marked advantages. The necessary paths are generated as required and need not be kept from one iteration to the next. Thus the storage requirements are modest and do not increase with the number of iterations. At each iteration only the arc flows v_a^ℓ need to be kept, in addition to the link costs $s_a(v_a^\ell)$ and the network. A drawback of this algorithm is that it converges slowly in the vicinity of the optimal solution. In practice this is not very limiting, since very accurate solutions are not needed for large scale problems.

The variable demand problem is just slightly more difficult to solve than the fixed demand problem. From a pedagogical viewpoint the following transformation of the variable demand problem into a fixed demand problem is instructive (see Dantzig *et al.*[10]). For each O/D pair i a node and two arcs are added as indicated below



For the pair i the origin node is p and the destination node is q. A node q' and the arcs (p,q') and (q,q') are added. The cost of arc (p,q') is zero and the user cost function on arc (q,q') is $-W_i(g_i)$. A demand of g_i^{max}, which is an upper bound on the equilibrium demand is imposed from p to q. The solution of this fixed demand problem has the property that all used paths are of length 0; thus any path on the original network that carries flow has length u_i^* and therefore $u_i^*=W_i(g_i^*)$ or $g_i^* = G_i(u_i^*)$.

This transformation is not efficient computationally since the arc (p,q') is not necessary in the straightforward adaptation of the linear approximation method to solve the variable demand problem, which only requires the addition of the arc (q,q'), or for all O/D pairs, the addition of a matrix of dimensions $|I| \times |I|$ which

contains the current O/D matrix at each iteration. For the variable demand network equilibrium problem the linearized problem at iteration ℓ is

$$\text{Min} \sum_{a\epsilon A} s_a(v_a^\ell) \sum_{i\epsilon I} \sum_{k\epsilon K_i} \delta_{ak} \cdot f_k - \sum_{i\epsilon I} W_i(g_i^\ell)\, w_i \tag{34}$$

subject to

$$\sum_{k\epsilon K_i} f_k - w_i = 0, \qquad i\epsilon I \tag{35}$$

$$f_k \geq 0, \quad k\epsilon K_i, \; i\epsilon I, \; w_i \geq 0, \quad i\epsilon I \tag{36}$$

As before, the objective (34) simplifies after some manipulations to

$$\text{Min} \sum_{i\epsilon I} \left(\sum_{k\epsilon K_i} s_k^\ell f_k - W_i(g_i^\ell) w_i \right) \tag{37}$$

which may be solved by determining u_i to be lengths of the shortest paths based on the current link cost $s_a(v_a^\ell)$ and determining the demand to be allocated to these shortest path as follows, where g_i^{max} is an upper bound of g_i.

$$w_i^\ell = \begin{cases} g_i^{max} & \text{if} \quad u_i < W_i(g_i^\ell) \\ g_i^\ell & \text{if} \quad |u_i - W_i(g_i^\ell)| < \varepsilon \quad i\epsilon I \\ 0 & \text{if} \quad u_i > W_i(g_i^\ell) \end{cases} \tag{38}$$

The descent direction is $\{(f_a^\ell - v_a^\ell), (w_i^\ell - g_i^\ell)\}$ where f_a^ℓ is obtained by assigning the demand w_i^ℓ to the shortest paths. The step size is determined by finding λ^ℓ which annuls the gradient

$$\sum_{a\epsilon A} s_a(v_a^\ell + \lambda(f_a^\ell - v_a^\ell))(f_a^\ell - v_a^\ell) - \sum_{i\epsilon I} W_i(g_i^\ell + \lambda(w_i^\ell - g_i^\ell))(w_i^\ell - g_i^\ell) \tag{39}$$

An interesting variant of the linear approximation method for solving the variable demand network equilibrium problem is that of partial linear approximation that was first suggested by S. Evans [14] for solving the combined trip distribution/assignment problem. This problem is a variant of the variable demand problem where the demand functions are taken to be negative exponential function of the origin to destination travel cost, that is $G_i(u_i) \cong \exp(-\theta u_i)$ and additional constraints are imposed on the total trips originating at origins p and terminating at destinations q. This model may be stated as

$$\text{Min} \sum_{a\epsilon A} \int_0^{v_a} s_a(x)dx + \theta \sum_{i\epsilon I} g_i \; \ell n \; g_i \tag{40}$$

subject to

$$\sum_{k\epsilon K_i} h_k - g_i = 0 \;, \quad i\epsilon I \tag{41}$$

$$\left.\begin{aligned} \sum_p g_{pq} &= D_q, \quad \forall q \\ \sum_q g_{pq} &= O_p, \quad \forall p \end{aligned}\right\} i=(p,q) \tag{42}$$

$$h_k \geq 0, \quad k\epsilon K_i, \; i\epsilon I; \quad g_i \geq 0, \; i\epsilon I \tag{43}$$

See also [17]. Although well studied, this model has not yet found application in practice.

Evans' algorithm was interpreted by Stewart [33] to be a partial linear approximation, where, for the variable demand problem the user cost functions are linearized, but the inverse demand functions are not. The resulting subproblem is

$$\text{Min} \sum_{a\epsilon A} s_a(v_a^\ell) \sum_{i\epsilon I} \sum_{k\epsilon K_i} \delta_{ak} \cdot f_k - \sum_{i\epsilon I} W_i(w_i) \tag{44}$$

$$\text{subject to} \quad \sum_{k\epsilon K_i} f_k - w_i = 0, \quad i\epsilon I \tag{45}$$

$$f_k \geq 0, \; k\epsilon K_i, \; i\epsilon I; \quad w_i \geq 0, \; i\epsilon I \tag{46}$$

This subproblem is still easy to solve by determining u_i to be the shortest paths based on the current link cost $s_a(v_a^\ell)$ and then determining the demand to be allocated to these shortest paths as follows

$$w_i^\ell = G_i(u_i), \quad i\epsilon I \tag{47}$$

The partial linear approximation direction may be easily shown to be a descent direction for the general problem Min $f_1(x) + f_2(y)$ subject to $A_x x + A_y y = b$, $x,y \geq 0$ by considering the optimality conditions for the subproblem Min $\nabla f_1(x^\ell)x + f_2(y)$ which are $\nabla f_1(x^\ell)(x-\tilde{x}) + \nabla f_2(\tilde{y})(y-\tilde{y}) \geq 0$, where $\sim$ denotes the optimal solution of the subproblem. Now let $x = x^\ell$ and by using the property that the gradient of a convex function is monotone, one obtains that $\nabla f_1(x^\ell)(\tilde{x}-x^\ell) + \nabla f_2(y^\ell)(\tilde{y}-y^\ell) \leq 0$. Thus, unless the current solution x^ℓ is optimal, in which case $\nabla f_1(x^\ell)(\tilde{x}-x^\ell) + \nabla f_2(y^\ell)(\tilde{y}-y^\ell) \geq 0$, a direction of descent is obtained. (Nguyen provided first a different proof of the above, as reported in [33])

Intuitively, this direction should be better, since the subproblem used for finding the direction is "closer" to the original problem. Computational results [19,23] support this intuition. As this subproblem is not more difficult to solve than the linear approximation subproblems, it is attractive to use in practice.

Another variant of the linear approximation method which is particularly useful for the variable demand version of this problem is the parallel tangent or PARTAN variant. At each step of the linear approximation algorithm another direction of descent is generated by every other solution $v^{\ell-2}$ and v^ℓ. Thus the new solution is generated by finding α^ℓ, $\alpha^\ell \leq \bar{\alpha}$ which minimizes the objective function for the solution $v^{\ell-2} + \alpha(v^\ell - v^{\ell-2})$, where $\bar{\alpha}$ is the largest step size allowed in order to maintain the nonnegativity of the path flows. Note that this direction is not guaranteed to be a descent

direction and, in particular if $\alpha^{\ell} < 1$, the objective function decreases in the opposite direction. The implementation of this variant requires the storage of all the weights that are given to each solution of the direction finding subproblem, of the step sizes α^{ℓ} taken in the PARTAN direction and of the solutions $v^{\ell-1}$ and $v^{\ell-2}$. The computational results obtained with this method suggest that it is a worthwhile modification of the basic method. Its additional storage requirements are modest for the fixed demand network equilibrium problem and for problems of moderate size, it may be suitable for the variable demand problem as well.

SYSTEM OPTIMAL vs USER OPTIMAL FLOWS

Wardrop's second principle is usually refered to as yielding "system optimal" flows since these are the flows which minimize the total cost in the system. The fixed demand version of this problem is

$$\text{Min} \sum_{a \epsilon A} v_a \cdot s_a(v_a) \tag{48}$$

subject to (23), (24).

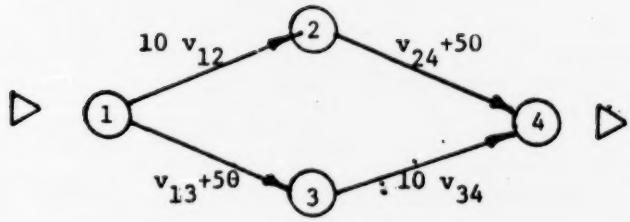
By applying the Kuhn-Tucker conditions to the system optimal problem one obtains that the system optimal flows have equal marginal cost on all the used paths, that is

$$m_k = \sum_a \delta_{ak} \left(s_a(v_a) + v_a \cdot s_a'(v_a)\right) \tag{49}$$

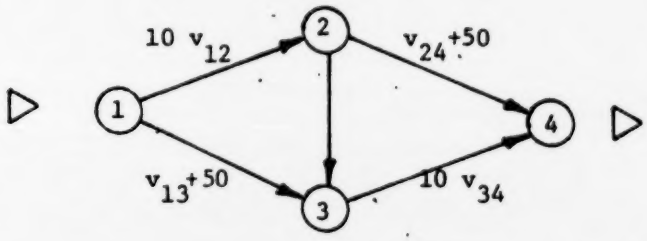
$$\text{and} \quad m_k^* - u_i^* \begin{cases} = 0 & \text{if} \quad h_k^* > 0 \\ \geq 0 & \text{if} \quad h_k^* = 0, \; k \epsilon K_i, \; i \epsilon I \end{cases} \tag{50}$$

The computation of the system optimal flows may be carried out by the linear approximation algorithm by making the appropriate modification of objective function and use of marginal link costs in the direction finding subproblem. Since the system optimal flows are more "efficient", in the sense that the trips are undertaken with a minimal total cost, than the user optimal flows, if link tolls equal to $v_a \cdot s_a'(v_a)$ could be imposed, then the users seeking their best routes will distribute themselves over the paths of the network in a system optimal way. The practical difficulty with imposing such tolls is that any tolls are likely to change the original user cost functions as well. Economists refer to such tolls as marginal cost pricing. The notion of marginal cost pricing may be found as early as 1844 with the work of Dupuit [13]. See also [27].

In general, the user optimal flows are different then the system optimal flows in a given network. For large, uncongested networks the difference may be small, since for constant link travel cost there is no difference between the user and system optimal flows. However these differences are significant when evaluating different scenarios where link additions or deletions are considered. Braess [3] (see also Murchland [26]) provided a classical example of such differences, by using the network shown below where the user cost functions are indicated on the links. The equilibrium flows are 3 units on the



path (1,2)(2,4) and 3 units on the path (1,3)(3,4). The total cost is 6 x 83 = 498. If the link (2,3) is now added to the network with user cost function $v_{23} + 10$



the equilibrium flows are 2 units on each of the paths of the network with total cost of 6 x 92 = 552. Thus, the addition of an arc increases the total travel cost as well as the travel cost to each user. Similarly, the suppression of the arc (2,3) would reduce the total travel cost and the cost to each user. This apparent paradox is understood by examining the system optimal solution which is 3 units each along the paths (1,2)(2,4) and (1,3)(3,4), even when the arc (2,3) is present. There is no reason to anticipate that total travel cost may be reduced when a link is added, since these costs are not minimized by the user optimal flows.

The occurrence of the Braess' paradox is of significant importance in the decisions taken regarding the additions or deletions of links in a network. If the user optimal principle is suitable as the behavioral representation of the route choice, then given a particular network, it may be advantageous to restrict the use of some of its links and care must be exercised to properly evaluate link additions to detect possible occurrences of increase in total costs and increase of service levels for some of the users. The Braess' paradox may certainly occur in real networks, as evidenced by the work of Fisk and Pallottino [15].

The other relevant question related to system optimal flows is the form of the problem with variable demand. One appropriate objective function for this model is

$$\text{Min} \sum_{a\epsilon A} v_a \cdot s_a(v_a) - \sum_{i\epsilon I} \int_0^{g_i} W_i(y)dy \qquad (51)$$

subject to (2), (3).

By applying the Kuhn-Tucker conditions to this problem one finds (50) and

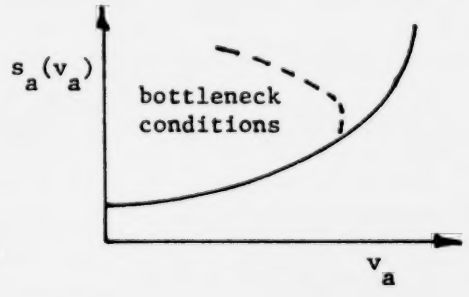
$$\left.\begin{aligned} G_i(u_i^*) &= g_i^* \\ \text{where} \quad u_i^* &= \min_{k\epsilon K_i} m_k^* \end{aligned}\right\} i\epsilon I \qquad (52)$$

that is, the cost perceived by the user is the marginal cost, rather than the average cost. This model is suitable for applications in rail networks where it may be argued that the tariffs are proportional to the marginal costs.

APPLICATION AND VALIDATION OF THE NETWORK EQUILIBRIUM MODEL

The validation of the network equilibrium model is based at present on relatively scarce empirical studies. Florian and Nguyen [16] applied the model to the urban network of the City of Winnipeg and Dow and Van Vliet [12] applied it to the City of Leeds. A practical problem with applying the network equilibrium model is the determination of the origin/destination matrix. Even when survey data is available, it does not include information for all the trips taken during the peak hour and the flow predictions must be reconciled with observed link counts. The process of calibrating a network equilibrium model involves the network representation, the calibration and assignment of user cost functions to the links of the network and possible adjustments of the origin/destination matrix. It may be stated that some significant approximations are often made in order to build the necessary data base for the application of the model. Yet, it may be the best predictive tool available today for predicting the impact of network changes in the short and medium term. The link flows obtained in the studies [16,12] simulate flows on congested links quite well and the origin to destination travel times are satisfactorily reproduced as well.

The application of the network equilibrium model is based on the assumption that the traffic is under free flow conditions and that no severe bottlenecks occur. When this assumption is not satisfied, increasing user cost functions do not model properly the phenomenon of increased travel time and reduced flow on links subject to bottlenecks, as indicated in the graph below.



The use of the network equilibrium model is not particularly expensive in computer expenditures. Experience with network of medium size (200 origins/destinations, 1500 nodes and 4000 links) indicate that, on a typical mainframe computer, the cost of 10-15 iterations of the linear approximation method is in the range of $150 - $300. The possibility of carrying out these computations on the new generation of 16 bit micro-computers will probably render the cost even lower.

The network equilibrium model may also be used to measure fuel consumption impacts by incorporating the empirical fuel consumption equations of Chang *et al.* [5] in an appropriate way.

MULTIPATH FLOWS WITH CONSTANT LINK COSTS

The user equilibrium flows are suitable for representing route choice behavior when congestion effects significantly increase travel costs. When the link user costs are constant, the user optimal principle is satisfied by assigning all the flows to the shortest cost path between origines and destinations (recall (25)-(27) and (28), (29)). Since it is seldom in practice that travellers have exact knowledge of the link times, it is common that more than one route is used, even when no congestion is present. Dial [11] proposed a route choice model based on the logit model which postulates that

$$h_{k'} = \bar{g}_i \cdot \exp(-\theta s_{k'}) / \sum_{k \in K_i} \exp(-\theta s_k) \qquad (53)$$

subject to the usual flow conservation and nonnegativity constraints (23), (24). This model has the equivalent optimization problem

$$\text{Min} \sum_{i \in I} \sum_{k \in K_i} h_k \, (\ell n \, h_k + \theta s_k) \qquad (54)$$

subject to (23), (24). The solution of this model requires the explicit storage of the paths. Dial ingeniously circumvented this problem by restricting the paths to be considered and succeeded in solving this restricted version of the problem without requiring path enumeration.

Although this model found much use in practice, the use of the logit path choice makes the resulting flows very sensitive to the network representation and the restriction of the paths results in the model being unstable; slight modifications in the network may induce large changes in the resulting flows.

COMMENTARY

The network models that are used in transportation planning have undergone rapid development in the past decade. Most of the material presented in these lectures is not included in the highly regarded book by Potts and Oliver [31] published in 1972. It is reasonable to anticipate that the applications of these models in various contexts will increase in the coming years.

REFERENCES

1. H.Z. Aashtiani and T.L. Magnanti, "Equilibria on a Congested Transportation Network". Opns. Res. Ctr., M.I.T., 1980. To appear in SIAM Journal on Algebraic and Discrete Methods.

2. M.J. Beckmann, C.B. McGuire and C.B. Winsten, "Studies in the Economics of Transportation". Yale University Press, New Haven, CT, 1956.

3. D. Braess, "Uber ein Paradoxen der Verkehrsplanning". Untermehmens forschung, pp. 258-268, 1968.

4. M. Bruynooghe, A. Gibert and M. Sakarovitch, "Une méthode d'affectation du trafic". In Fourth Symposium on the Theory of Traffic Flow, Karlsruhe, 1968.

5. M.F. Chang, L. Evans, R. Herman and P. Wasiliewski, "Gasoline Consumption in Urban Traffic." Transportation Research Record 599, pp. 25-30, 1976.

6. S. Dafermos and A. Nagurney, "Sensitivity Analysis for the General Asymmetric Network Equilibrium Problem." Technical Report LCDS # M-82-7, Lefschetz Center for Dynamical Systems, Brown University, 1982.

7. S. Dafermos, "Traffic Equilibrium and Variational Inequalities." Transportation Science, Vol. 14, pp. 42-54, 1980.

8. S.C. Dafermos, "The General Multimodal Network Equilibrium Problem with Elastic Demand." Lefaschetz Center for Dynamical Systems, Brown University, 1980. To appear in Networks.

9. S. Dafermos, "An Extended Traffic Assignment Model with Applications to Two-way Traffic." Transportation Science, 4, pp. 336-389, 1971.

10. G . Dantzig, S.F. Maier and Z.F. Lansdowne, "Application of Decomposition to Transportation Network Analysis." Tech. Report DOT-TSC-OST-76-26, Control Analysis Corporation, 1976.

11. R.B. Dial, "A Probabilistic Multipath Assignment Model which Obviates Path Enumeration." Transportation Research, 8, pp. 85-96, 1971.

12. P. Dow and D. Van Vliet, "Capacity Restrained Road Assignment." Traffic Engineering and Control, pp. 261-273, 1979.

13. J. Dupuit, "De la mesure de l'utilité des travaux publics." Annales des travaux publics, Annales des ponts et chaussées, 2e série, 8, 1844.

14. S.P. Evans, "Derivation and Analysis of Some Models for Combining Trip Distribution and Assignment." Transportation Research, 10, pp. 37-57, 1976.

15. C. Fisk and S. Pallottino, "Empirical Evidence for Equilibrium Paradoxes with Implications for Optimal Planning Strategies." Transportation Research A, Vol. 15A, No. 3, pp. 245-248, 1981.

16. M. Florian and S. Nguyen, "An Application and Validation of Equilibrium Trip Assignment Methods." Transportation Science, Vol. 10, pp. 374-389, 1976.

17. M. Florian, J. Ferland and S. Nguyen, "On the Combined Distribution-Assignment of Traffic." Transportation Science, 9, pp. 43-53, 1975.

18. M. Frank and P. Wolfe, "An Algorithm of Quadratic Programming." Naval Res. Log. Quart. 3, pp. 95-110, 1956.

19. C. Frank, "A Study of Alternative Approaches to Combined Trip Distribution-Assignment Modeling." Ph.D. Thesis, Department of Regional Science, University of Pennsylvania, 1978.

20. M. Hall, "Properties of the Equilibrium State in Transportation Networks." Transportation Science, Vol. 12, pp. 208-216, 1978.

21. D. Kinderlehrer and G. Stampacchia, "An Introduction to Variational Inequalities and Applications." Academic Press, 1980.

22. F.H. Knight, "Some Fallacies in the Interpretation of Social Cost." Quarterly Journal of Economics, 38, pp. 582-606, 1924.

23. L.J. LeBlanc and K. Farhangian, "Efficient Algorithms for Solving Elastic Demand Traffic Assignment Problems and Mode Split Assignment Problems." Transportation Science, 15, pp. 306-312, 1981.

24. L.J. LeBlanc, E.K. Morlok and W.P. Pierskalla, "An Efficient Approach to Solving the Road Network Equilibrium Traffic Assignment

Problem." Transportation Research 5, pp. 309-318, 1975.

25. L.J. LeBlanc, "Mathematical Programming Algorithms for Large Scale Network Equilibrium and Network Design Problems." Ph.D. Dissertation, Dept. of Ind. Eng. and Mgt. Sci., Northwestern University, 1973.

26. J.D. Murchland, "Braess' Paradox of Traffic Flow", Transportation Research, 4, pp. 391-394, 1970.

27. M. Netter, "Equilibrium and Marginal Cost Pricing on a Road Network with Several Traffic Types." In Traffic Flow and Transportation, Proceedings of the Fifth International Symposium on the Theory and Traffic Flow and Transportation, G.F. Newell (ed.), Springer-Verlag, 1971.

28. S. Nguyen, "Procedures for Equilibrium Traffic Assignment with Elastic Demand." Publication #39, Centre de recherche sur les transports, Université de Montréal, 1977.

29. S. Nguyen, "A Unified Approach to Equilibrium Methods for Traffic Assignment." In Traffic Equilibrium Methods, Proceedings 1974 (M. Florian, Editor), Vol. 118 Lecture Notes in Economics and Mathematical Systems, Springer Verlag, pp. 148-182, 1976.

30. S. Nguyen "Une approche unifiée des méthodes d'équilibre pour l'affectation du trafic." Ph.D. Dissertation, Département d'informatique et de recherche opérationnelle, Université de Montréal, 1973.

31. R.B. Potts and R.M. Oliver, "Flows on Transportation Networks." Academic Press, New York, 1972.

32. M.J. Smith, "Existence, Uniqueness and Stability of Traffic Equilibria." Transportation Research B, Vol. 13B, pp. 295-304, 1979.

33. N.F. Stewart, "Notes on the Mathematical Structure of Equilibrium Models." Technical Report LITH-MATH-12-79-8, Department of Mathematics, Linkoping Institute of Technology, 1979.

34. J.G. Wardrop, "Some Theoretical Aspects of Road Traffic Research." Proc. Inst. Civil Engineers, Part II, pp. 325-378, 1952.

APPENDIX I - CONVEX AND MONOTONE FUNCTIONS

DEFINITION-A numerical function f defined over a convex set Ω is *convex* if, for every $x, y \in \Omega$ and for α, $0 \leq \alpha \leq 1$

$$\alpha f(x) + (1-\alpha) f(y) \geq f(\alpha x + (1-\alpha)y)$$

If the above inequality is satisfied as a strict inequality for $x \neq y, \alpha$, $0 < \alpha < 1$, then the function f is *strictly convex*.

DEFINITION-A numerical function g defined on a convex set Ω is said to be *concave* if the function $f = -g$ is convex. The function g is *strictly concave* if -g is strictly convex.

PROPERTY 1-Let f_1 and f_2 be convex functions defined over a convex set Ω. Then the function $f_1 + f_2$ is convex on Ω.

PROPERTY 2-Let f be a convex function over the convex set Ω. The function af is convex for any $a \geq 0$.

There are several important and useful properties of differentiable convex functions.

PROPERTY 3-Let f be a differentiable convex function defined on a convex set Ω. f is convex if and only if $f(y) \geq f(x) + \nabla f(x)(y-x)$ for all $x, y \in \Omega$.

DEFINITION-A vector function h defined on a convex set Ω is *monotone* if $(h(x)-h(y))(x-y) \geq 0$ for $x, y \in \Omega$. If the above inequality is satisfied as a strict inequality for $x \neq y$, $x, y \in \Omega$, then the function is *strictly monotone*. If a function h is monotone its negative is *antitone*.

PROPERTY 4-Let f be a differentiable convex function defined on a convex set Ω. f is convex if and only if ∇f is monotone, that is $(\nabla f(x)-\nabla f(y))(x-y) \geq 0$ for $x, y \in \Omega$.

The following are classical results concerning the minimization of convex functions.

PROPERTY 5- Let f be a convex function defined on a convex set Ω. Then f achieves its minimum over a convex set Γ and any local minimum of f is a global minimum.

PROPERTY 6-Let f be a differentiable convex function defined on a convex set Ω. x^* is a global minimum of f over Ω if for all $y \in \Omega$, $\nabla f(x^*)(y-x^*) \geq 0$.

REFERENCES FOR APPENDIX I

- O.L. Mangasarian, "Nonlinear Programming." McGraw Hill, 1964.
- D.G. Luenberger, "Introduction to Linear and Nonlinear Programming." Addison-Wesley, 1965.

APPENDIX II - PRIMAL METHODS FOR NONLINEAR PROGRAMS

Consider the solution of the nonlinear programming problem

$$\begin{aligned} &\min f(x) \\ &\text{subject to} \quad g(x) \leq 0 \\ &\qquad\qquad\ \ h(x) = 0 \end{aligned}$$

A *primal method* of solution is a procedure that works on the problem directly by searching through the feasible region for the optimal solution. Each point generated by the procedure is feasible and the value of the objective function decreases constantly.

A *feasible direction* d_k is a vector that permits, given a feasible x_k^ℓ to generate $x^{\ell+1} = x^{k\ell} + \lambda^\ell d^\ell$, where λ^ℓ is a nonnegative scalar. This scalar is chosen to minimize the objective function f such that x_{k+1} is feasible. Thus a vector d_k is a feasible direction at x_k if there is a $\bar{\lambda} > 0$ such that $x_k + \lambda d_k$ is feasible for all λ, $0 \leq \lambda \leq \bar{\lambda}$ and $\nabla f(x^\ell)^T d^\ell \leq 0$.

One of the simplest feasible direction methods for the problem of minimizing a convex function f subject to the linear constraints $Ax = b$, $x \geq 0$, is the linear approximation method. Given a feasible point x_k, the direction vector is chosen to be $(y^\ell - x^\ell)$ where y^ℓ is the solution of the linear program

$$\begin{aligned} &\min \nabla f(x^\ell)^T y \\ &\text{subject to} \quad Ay = b \\ &\qquad\qquad\ \ y \geq 0 \end{aligned}$$

$d^\ell = (y^\ell - x^\ell)$ is a direction of descent since the optimality conditions of the above linear program are that $\nabla f(x^\ell)^T(y-y^\ell) \geq 0$ and for $y = x^\ell$ it follows that $\nabla f(x^\ell)^T(y^\ell - x^\ell) < 0$ unless $x^\ell = x^*$, the solution of the convex program, where $\nabla f(x^*)^T(y-x^*) \geq 0$.

The detailed steps of the linear approximation algorithm for solving this problem are :

STEP 1-Given a feasible solution x^1, set $\ell = 1$.

STEP 2-Determine y^ℓ that minimizes

$$\nabla f(x^\ell)y, \text{ subject to } A\,y = b,\ y \geq 0.$$

STEP 3-Set the descent direction $d^\ell = y^\ell - x^\ell$. If $|\nabla f(x^\ell)d^\ell| \leq \varepsilon$, terminate, where ε is a suitable convergence parameter, (x^ℓ is the optimal solution).

STEP 4-Find the optimal step length λ^ℓ that minimizes

$$F(x^\ell + \lambda d^\ell) \quad \text{for} \quad 0 \leq \lambda \leq 1.$$

STEP 5-Revise the current solution

$$x^{\ell+1} = x^\ell + \lambda^\ell d^\ell;$$

set $\ell = \ell + 1$ and return to Step 2.

APPENDIX III - THE KUHN-TUCKER CONDITIONS

Consider the nonlinear programming problem

$$\text{Min } f(x)$$

$$\text{subject to} \quad g(x) \leq 0$$
$$h(x) = 0$$

Let x^* be a feasible solution which is a local minimum of the above problem. Then there exists a vector λ and a vector μ, $\mu \geq 0$, such that

$$\nabla f(x^*) + \lambda \nabla h(x^*) + \mu \nabla g(x^*) = 0$$

$$\text{and} \qquad \mu g(x^*) = 0 .$$

If $f(x)$ is convex and the feasible set is convex as well, the Kuhn-Tucker conditions are necessary and sufficient for x^* to be a global minimum of $f(x)$.

APPENDIX IV - DUALITY THEORY FOR LINEAR PROGRAMS

Consider the linear program

$$\text{Min } cx$$

$$\text{subject to} \quad Ax = b$$
$$x \geq 0$$

and its corresponding dual problem

$$\text{Max } ub$$

$$\text{subject to} \quad uA \leq c$$

PROPERTY 1-If x and u are feasible primal and dual solution and $cx = ub$, then they are optimal solution as well.

PROPERTY 2- If x and u are feasible solutions for the primal and dual problems, a necessary and sufficient condition that they both be optimal is that $(uA-c)x = 0$.

REFERENCE FOR APPENDICES II, III and IV

- D.G. Luenberger, "Introduction to Linear and Nonlinear Programming." Addison-Wesley, 1965.